

# Weak equivalence principle for self-gravitating bodies: A sieve for purely metric theories of gravity

Eolo Di Casola\* and Stefano Liberati†  
*SISSA, via Bonomea 265, 34136 Trieste (Italy) and*  
*INFN Sez. di Trieste, via Valerio 2, 34127 Trieste (Italy)*

Sebastiano Sonego‡  
*DCFA, Sezione di Fisica e Matematica, Università di Udine, Via delle Scienze 206, 33100 Udine (Italy)*  
 (Dated: January 2014)

We propose the almost-geodesic motion of self-gravitating test bodies as a possible selection rule among metric theories of gravity. Starting from a heuristic statement, the “gravitational weak equivalence principle”, we build a formal, operative test able to probe the validity of the principle for any metric theory of gravity, in an arbitrary number of spacetime dimensions. We show that, if the theory admits a well-posed variational formulation, this test singles out only the purely metric theories of gravity. This conclusion reproduces known results in the cases of general relativity (as well as with a cosmological constant term), and scalar-tensor theories, but extends also to debated or unknown scenarios, such as the  $f(R)$  and Lanczos–Lovelock theories. We thus provide new tools going beyond the standard methods, where the latter turn out to be inconclusive or inapplicable.

PACS numbers: 04.20.Cv, 04.50.Kd, 04.50.-h, 04.40.-b.

Keywords: Equivalence principles, extended theories of gravity, self-gravitating bodies.

## I. INTRODUCTION

Despite the pivotal role played at the dawn of general relativity, the “principle of equivalence” is now regarded as a rather vague, heuristic statement, perhaps useful for teaching purposes, but surely unable to challenge the contemporary approach to gravity, based on sharp and neat mathematical axioms and precision-test experiments [1, 2]. Yet, the original idea behind the statement was nothing but brilliant: to provide a direct connection between the basic structural elements of a theory and intuitive elementary experiments [3].

Is it possible to reconcile the two points of view, building a rigorous model out of a crisp, yet somewhat imperfect, premise? We believe so, and we try to prove it by showing, in the present work, how an aptly refined version of the principle of equivalence can be smoothly translated into formal statements without shadowing the underlying physical intuition. We then make use of our construction to establish a few selection criteria in the arena of extended theories of gravitation, singling out a specific sub-class of frameworks.

To this end, we must first and foremost notice that “equivalence principle” is way too generic an expression. It is actually possible to distinguish several inequivalent formulations of the principle, each one identifying a different class of theories, with different fundamental mathematical structures. These principles are sometimes formulated rather loosely, and different authors present them in ways that are not exactly equivalent. There

is, however, general agreement about the formal implications and physical content of some of them.<sup>1</sup> Broadly speaking, the universality of free fall for test bodies with negligible self-gravity (*weak equivalence principle*, or WEP) implies that gravity can be described in terms of a nonflat connection  $\Gamma$  on spacetime.

The eliminability of gravity in the local description of fundamental, nongravitational phenomena (*Einstein’s equivalence principle*, or EEP), together with the generally accepted Minkowskian character of spacetime in the absence of gravity, implies that  $\Gamma$  is the Levi-Civita connection of a metric  $g_{ab}$  with a Lorentzian signature, thus singling out the so-called *metric theories of gravity*.<sup>2</sup> Moreover, within such theories, EEP provides a prescription for writing down physical laws in a gravitational field once they are known in Minkowski spacetime.

In the literature, one often finds yet another equivalence principle: the so-called *strong equivalence principle* (SEP), which extends the validity of EEP to gravitational phenomena as well; i.e., it postulates that a background gravitational field does not influence the outcome of local experiments involving gravity. The SEP implies of course EEP, but the converse is not true, so it is legitimate to ask which further restrictions does the SEP impose upon a metric theory of gravity.

<sup>1</sup> The review offered here of the various equivalence principles is inevitably limited and incomplete, as we highlight only those aspects that are relevant for the purpose of the present paper. See Ref. [3] for a more detailed examination of the various elements entering into the formulation of each of these principles.

<sup>2</sup> This is the very minimum requirement, and nothing forbids the presence of other gravitational degrees of freedom (scalars, vectors, tensors, etc.) in addition to the metric.

\* Corresponding author. E-mail address: eolo.dicasola@sissa.it

† E-mail address: stefano.liberati@sissa.it

‡ E-mail address: sebastiano.sonego@uniud.it

In this respect, it is useful to introduce also a somewhat intermediate step, in the form of a *gravitational weak equivalence principle* (GWEP). The latter extends the universality of free fall to self-gravitating test bodies.<sup>3</sup> The GWEP is clearly a necessary condition for the SEP (actually, its main founding pillar), but is not implied by EEP, as the latter deals exclusively with non-gravitational phenomena. Moreover, the validity of the GWEP is in many ways easier to establish than the one of the SEP, being linked, as we shall see, to clearer and more familiar geometrical statements. For all these reasons, we deem it appropriate to focus first and foremost on the GWEP itself, its formal implications, and operational implementation.

In four spacetime dimensions, the GWEP turns out to be an extremely selective requirement. Only the theories of Einstein and Nordström abide by such principle.<sup>4</sup> It is therefore interesting to know what characterises these theories and makes them “special”. Also, it is desirable to export — when possible — the results to the vast landscape of extended theories of gravity, and to higher-dimensional spacetimes (where there is a wide range of “natural extensions” of general relativity, the so-called Lanczos–Lovelock theories).

We start by discussing the formulation of the GWEP in Sec. II. Differently from what happens for the WEP, even in the case of a “minimal” theory like general relativity, it is not obvious at first sight whether the GWEP holds or not. Thus, in Sec. III, we develop a method, encompassing a wide class of gravitational theories (and in an arbitrary number of spacetime dimensions), to check the validity of the GWEP for a given framework. In Sec. IV, the method is applied to the specific cases of general relativity (as well as in the presence of a cosmological constant), scalar-tensor and higher-order theories, and Lanczos–Lovelock/Gauß–Bonnet theories. Our conclusions are presented in Sec. V.

Hereinafter, spacetime is modelled by a smooth  $m$ -dimensional Lorentzian manifold  $(\mathcal{M}, g_{ab})$ . Physical fields are geometric objects (tensorial or spinorial) on  $\mathcal{M}$ , and are all supposed to be dynamical — an assumption which enforces invariance under diffeomorphisms. This framework is wide enough to cover all the metric theories of gravity. We adopt the sign conventions of Wald’s textbook [6], and use units in which the value of the speed of light in vacuum is equal to 1.

## II. GRAVITATIONAL WEAK EQUIVALENCE PRINCIPLE

This section is devoted to reviewing the various formulations of the GWEP. We begin by proposing a heuristic, yet physically expressive, statement of this principle, and then discuss its relation with the equality between inertial and gravitational mass. Finally, we comment on a recent proposal for a condition that should be fulfilled by a theory in which the GWEP/SEP holds true.

### A. Intuitive formulation

The simplest, and most naïve, formulation of the GWEP can be stated as follows [1, 2]: *Freely-falling test bodies behave independently of their properties.*

Of course, this statement includes the WEP, which corresponds to the limit of negligible self-gravity. Note that, although we allow for a sizeable self-gravity, we retain the condition that we are dealing with a test body. The reason is that, if one considers also bodies whose back-reaction on the environment is non-negligible, it becomes impossible to compare their behaviour and then establish whether it is, or it is not, the same. In other words, the very notion of a *universal* behaviour requires that the environment remain unchanged when different bodies are considered. By definition, this condition is accomplished provided that one works only with test bodies.

The test-body limit, however, seems hard to obtain if one takes self-gravity into account. Consider, for example, the free fall of a micro black hole in the gravitational field of the Earth. Of course, although the presence of the black hole does not affect the behaviour of the Earth (and, thus, the gravitational field on large scales), it does affect quite strongly the gravitational field in its surroundings. Therefore, establishing whether the behaviour of the black hole is the same as the one of a different body (e.g., a pebble) looks problematic, because one is comparing two situations that might differ greatly in the region where motion takes place. In other words, self-gravitating bodies can never be considered test bodies on small scales. The way out of this deadlock is to pick an aptly chosen region surrounding the body, which extends as far as the influence of the latter becomes negligible (negligible with respect to the required precision, of course). Then, one ignores all the details within such a region, so that the body is represented just by a world tube in a spacetime that is not affected by the body itself.

We could then try to reformulate the GWEP as the property that different bodies are represented, if they have the same initial conditions, by the same world tubes. Unfortunately, this is still a vague statement, and it is not difficult to see that it is very hard, if not impossible, to reformulate it in a more precise way. All difficulties disappear, however, if one focusses on very small bodies, whose world tubes can be approximated well enough by worldlines. For all these reasons, we shall adopt the

<sup>3</sup> Those of being “self-gravitating” and “test” are logically and physically independent properties. Whether a body has a sizeable self-gravity, is something which can be established considering the ratio between its gravitational self-energy and its rest energy. On the other hand, a test body is simply one not affecting appreciably its environment. This subtle, yet relevant difference is further examined in Ref. [3].

<sup>4</sup> The scalar theory of gravity proposed by Gunnar Nordström is a neglected gem of forgotten lore, which gave an impressive boost to the development of Einstein’s grand synthesis. See Refs. [4, 5] for a recollection of the genesis and features of this theory.

following intuitive formulation:

**Gravitational Weak Equivalence Principle.** *The world lines of small freely falling test bodies do not depend on the peculiar physical properties of the bodies themselves.*

Basically, this is just an extension of the WEP, where the restriction that the particle self-gravity be negligible has been removed. In particular, among the “physical properties” mentioned in the formulation of the GWEP, there can also be a parameter measuring the ratio between the gravitational self-energy and the rest energy of the body (sometimes referred to as the compactness). Since, in a metric theory of gravity, the world-lines of non-self-gravitating test particles are the geodesics of the background geometry, the GWEP requires that this is also the case for small bodies whose self-gravity cannot be neglected.

Contrary to a widespread belief, the GWEP does not hold strictly, even in general relativity [7, 8]. Indeed, particles equipped with spin or multipole moments do not follow geodesics, because their equation of motion contains terms expressing the coupling of these quantities with the spacetime curvature [9]. Therefore, although these particles do not feel forces of nongravitational origin, their behaviour is not universal. This happens even in the limit in which self-gravity can be neglected, so the violation of universality occurs already at the level of the WEP. Moreover, taking self-gravity into account, there is a further, usually minuscule, correction that makes a particle world-line non-geodesic — the so-called “tail term” [10, 11] — which describes the force on the body due to gravitational waves produced by the motion of the body itself at an earlier time, and then backscattered off the background curvature.<sup>5</sup>

Concerning these violations, we can see that those of the WEP are, after all, rather harmless, because they all follow from the internal structure of the bodies considered (a feature which survives, even in the point particle limit, in the form of quantities like the particle spin and multipole moments). One may still argue that the WEP would hold exactly for structureless particles, although this last notion is somewhat remote from the experimental realm. The situation is slightly less clear in the case of the tail term, for its origin is precisely the particle self-gravity, that is, the property which is at stake when it comes to test the GWEP. It should be noted, however, that upon taking that term into account, the task

of comparing the behaviour of different particles becomes hopeless, because any such comparison would demand the knowledge of the entire history of each body. Trying to establish whether two systems behave in the same way would then require control over their entire past, which is of course impossible. For this reason, and considering the practical irrelevance of the “tail force” in view of its extremely small magnitude, we shall ignore such contributions in the following analysis. Basically, this corresponds to a restricted version of the GWEP, where one’s attention is focussed on bodies whose self-radiative effects are negligible.

## B. Relation with other equivalence principles

As seen, the GWEP nicely fits in a heterogeneous group of equivalence principles. It is, then, worth pointing out some of the relations and logical implications among the various group members [3].

As an essential request, all viable theories of gravity ought to reduce, in some limit, to Newton’s theory.<sup>6</sup> At the same time, the basic principles of dynamics do not say anything about possible relations between the gravitational mass  $m_g$ , and the inertial mass  $m_i$ . Actually, these quantities express, in principle, two very different properties of a body. Experiments, however, show that they are proportional to each other through a universal coefficient [1, 2], and can therefore be set equal by a suitable choice of units. It is then fair to introduce the following additional equivalence principle, named after Newton:

**Newton’s Equivalence Principle (NEP).** *In the Newtonian limit, the inertial and gravitational masses of a particle are equal.*

This formulation makes it possible to test NEP also for theories other than Newton’s. What really matters, in fact, is the notion of Newtonian limit, for the identification of an inertial and a gravitational mass is in general unambiguous only in those conditions. Moreover, one can easily see that the WEP implies NEP, while the converse is not necessarily true.<sup>7</sup>

More to the point, the GWEP implies the WEP (and thus also NEP for all those bodies whose self-gravity is

<sup>5</sup> Such a term exists if one describes the particle motion with respect to the unaffected background. However, if one adds, to the background metric, a perturbation corresponding to the “tail”, it turns out that the particle follows a geodesic of this new metric. Nevertheless, this does not corroborate the GWEP, because the perturbation — and thus, the new metric — depends on the mass of the body. Such dependance once again violates the required universality.

<sup>6</sup> Usually, this happens whenever weak gravitational fields and slow motions are considered. Nonetheless, other admissible proposals exist; for instance, the TeVeS theory can lead to the modified Newtonian dynamics (MOND) model; see Ref. [12] for a review. In this latter case, however, the modifications emerge only for very small accelerations, so Newtonian dynamics still apply in a wide regime.

<sup>7</sup> The difference between the WEP and NEP is subtle, whence the frequent identification of the two statements. In general, however, the equivalence is not legitimate. Only as long as the two masses enter in the equations of motion through their ratio  $m_i/m_g$  alone — as in Newton’s theory — does NEP also imply the universality of free fall, and one has  $\text{WEP} \Leftrightarrow \text{NEP}$ .

negligible). Furthermore, the GWEP also implies NEP for all self-gravitating bodies. These last implications strongly suggest to further distinguish NEP from a *gravitational NEP* (GNEP) [3]. It follows, then, that every time the GNEP fails, the GWEP is violated as well. Given a theory of gravity in which the WEP holds, it is then possible to check whether NEP holds as well for a body with non-negligible self-gravity. The state-of-the-art technique to do so is the parametrized post-Newtonian (PPN) formalism [1, 2]. Among its main conclusions, we underline that, although in general relativity  $m_i$  always equals  $m_g$ , in the Brans–Dicke theory the two masses coincide only when self-gravity can be neglected [13].

That EEP implies the WEP is another straightforward conclusion. Since EEP guarantees that a freely-falling particle in a gravitational field behaves, locally, in a way indistinguishable from that of a free particle in the absence of gravity; and since the behaviour of free particles in the absence of gravity is universal, it follows that freely-falling particles also behave in a universal way. Similarly, one can see that the SEP implies the GWEP. The implication  $\text{WEP} \Rightarrow \text{EEP}$ , usually referred to as Schiff’s conjecture [1], is still debated.<sup>8</sup>

Finally, since the SEP is just EEP extended to “test gravitational phenomena” (of which the free fall of a test body with self-gravity is only a particular case), then it makes sense to ask whether, adding the GWEP to EEP, the full SEP could be recovered. To our knowledge, this “gravitational extension” of Schiff’s conjecture has never been advanced before but, if correct, it would single out the GWEP as the key element of the SEP.

### C. A recent approach to the SEP

Any principle is effective as a selection rule only as long as it can be operatively used to characterise different theories. The GWEP and the SEP provided so far still need significant improvement in this respect.

An interesting proposal attempting to better pin down the formal content of the SEP has been advanced in Refs. [14, 15]. It builds upon the existence of an analogy between general relativity and non-Abelian gauge theories *à la* Yang–Mills. One can then trade the usual dynamics of the metric for that of the connection (which plays the role of the gauge potential, while the curvature represents the field strength). The “true” field equations thus become

$$\nabla_d R_{abc}{}^d = \lambda j_{abc} . \quad (1)$$

These are second-order differential equations for the connection coefficients, sourced by “currents”  $j_{abc}$  ( $\lambda$  is a

coupling constant). Consequently, the Riemann curvature tensor is subject to the necessary condition that

$$\nabla_d R_{abc}{}^d = 0 \quad \text{in vacuum} . \quad (2)$$

Since the left-hand side can be cast, in view of the Bianchi identities, in the equivalent form

$$\nabla_d R_{abc}{}^d = \nabla_b R_{ac} - \nabla_a R_{bc} , \quad (3)$$

condition (2) is satisfied by general relativity. In Refs. [14, 15], it is regarded as a statement about the nonlinear character of the gravitational interaction — “gravitons gravitate as gluons glue”, in the words of Ref. [14]. Condition (2) is also claimed to imply the SEP because, imposing it as a set of constraints on an asymptotically flat, spherically symmetric metric representing a weak stationary field, one finds exactly the conditions on the PPN parameters that guarantee the validity of the NEP for self-gravitating bodies [13].

This proposal has for sure the merit of trying to turn the SEP into a formally precise statement. It must be stressed, however, that the PPN constraints emerging from condition (2) only provide a test of the GNEP for bodies with weak self-gravity [13]. Unless this implies the SEP, it is not obvious that Eq. (2) can be taken as a sufficient condition for the SEP.

Furthermore, while it is proven that Eq. (2) fails for the Brans–Dicke theory (as expected from a test of the SEP), an extension to other theories of gravity does not appear at all straightforward, for no general prescription is offered to build the equivalent of Eq. (2) in the case of, e.g., multi-scalar-tensor theories, higher-curvature theories and so forth.

For all these reasons, we shall try now to make a further step ahead, and propose a precise, independent formulation of the GWEP which can also be operatively useful.

### III. A CRITERION FOR THE GWEP

The formulation of the GWEP presented in Sec. II A offers a natural procedure for probing the validity of the principle for a metric theory of gravity. Given any test particle, treated as non-self-gravitating, we consider its world line on an assigned arbitrary background, and then try to prove that, when self-gravity is “turned on”, its new world-line coincides with the previous one.

Such criterion, it is worth stressing again, is intrinsically perturbative in nature, for one ends up checking that a small body — treated alternatively as non-self-gravitating and self-gravitating — will basically follow the same path on a given background spacetime. We are implicitly conjecturing, then, that the perturbation induced by “switching on” the self-gravitation of the body can be suitably neglected outside an appropriate world tube around the test body, or, stated otherwise, that the self-gravitating and the non-self-gravitating body live on the *same* background. This point is crucial, as it allows

<sup>8</sup> Interestingly enough, if Schiff’s conjecture proved to be true, a connection between the GWEP and EEP could be established, stemming from the implication  $\text{GWEP} \Rightarrow \text{WEP}$ .

to sidestep all the complications related to possible effects of the body on the surrounding geometry, and hence it relieves completely the ambiguities associated with the expression “behaving in the same way”.

The computation of the world-tube for a self-gravitating body is, in general, quite a difficult task. Nonetheless, as just stressed, all we need is an acid test of the fact that the world tube of a self-gravitating body contains a geodesic of the original background spacetime, rather than a detailed description of it. Such a test is viable, and is facilitated by a well-known theorem.

By appropriately deploying the content of such proposition, the perturbative approach intrinsic to the problem, and finally the variational formulation of a metric theory of gravity, it becomes possible to give the heuristic criterion a fully formal status, allowing for a straightforward check of its validity.

### A. Motion of a small body

In 1975, Geroch and Jang proved the following statement, refined some years later by Ehlers and Geroch [16, 17]:<sup>9</sup>

**Theorem (Geroch, Jang).** *Let  $\mathcal{C}$  be a smooth curve in a spacetime  $(\mathcal{M}, g_{ab})$ . Suppose that, given any open subset  $\mathcal{U}$  of  $\mathcal{M}$  containing  $\mathcal{C}$ , there exists a smooth symmetric field  $\Theta_{ab}$  on  $\mathcal{M}$  such that: (a)  $\Theta_{ab}$  satisfies the strengthened dominant energy condition;<sup>10</sup> (b)  $\Theta_{ab} \neq 0$  at some point in  $\mathcal{U}$ ; (c)  $\Theta_{ab} = 0$  outside of  $\mathcal{U}$ ; (d)  $\nabla^b \Theta_{ab} = 0$ . Then,  $\mathcal{C}$  is a timelike geodesic on  $(\mathcal{M}, g_{ab})$ .*

This seems to be exactly what we need to implement the universality of free fall. Indeed, the stress-energy-momentum tensor  $T_{ab}$  of a localised body satisfies conditions (a)–(d) above.<sup>11</sup> Hence, its world tube always contains a timelike geodesic, to which it can be approximated when the spatial extension of the body can be neglected.

Yet, this conclusion is not entirely satisfactory, because it is easy to see that conditions (a)–(d) hold independently of whether the body is, or is not, self-gravitating. The world-tube of a very small body can then always be approximated by a geodesic, and the GWEP seems to be satisfied trivially. Moreover, it appears that there is no substantial difference between the WEP and the GWEP, because they both hold under exactly the same

conditions. Such a conclusion looks suspicious, to say the least.

A little thinking shows, however, that, although it is true that the Geroch–Jang theorem implies geodesic motion for both a non-self-gravitating and a self-gravitating small body, such geodesics belong to *different* spacetimes. In the former case, the world tube approximates a geodesic in a background spacetime that does not satisfy the gravitational field equations with the body acting as a source. In the latter case, on the other hand, the world tube approximates a geodesic of the spacetime where the body is included among the sources of gravity — a spacetime varying from one body to another. Therefore, while the Geroch–Jang theorem suffices to prove the validity of the WEP, it does not imply that the GWEP also holds. As we pointed out in Sec. II A, for this purpose one ought to show that small self-gravitating bodies all have the same set of world lines in a background spacetime which is not affected by their presence. Since the stress-energy-momentum tensor  $T_{ab}$  satisfies condition (d) when  $\nabla$  is the covariant derivative with respect to the “full” spacetime, i.e., the one that solves the gravitational field equations considering the body itself as a source, but not with respect to the background, one cannot conclude that the universality of free fall automatically extends to self-gravitating bodies as well.

The GWEP would still hold, nevertheless, if one could find *another* symmetric tensor  $\Theta_{ab}$ , satisfying conditions (a)–(d) in the background. In particular, we stress that condition (d) should take the form

$$\bar{\nabla}^b \Theta_{ab} = 0, \quad (4)$$

where the covariant derivative  $\bar{\nabla}$ , and the metric  $\bar{g}_{ab}$  used to raise the index, now both refer to the background geometry. Of course, for a self-gravitating body, such a  $\Theta_{ab}$  cannot coincide with the stress-energy tensor  $T_{ab}$ . It is plausible that, if such a  $\Theta_{ab}$  exists at all, it should include also a contribution associated with self-gravitation.

As we shall see below, a perturbative approach naturally leads to a reasonable candidate for  $\Theta_{ab}$  — even better, one satisfying condition (4) in the case of general relativity, but not in the case of other theories that are known to violate the GWEP. In fact, whether the  $\Theta_{ab}$  emerging from the perturbative expansion does, or does not, satisfy condition (4), can be traced back to a specific property of the variational formulation of the theory of gravity under consideration. The tough problem of checking the validity of the GWEP is thus reduced to a much simpler one, namely, the mere inspection of the action functional.

Remarkably, the Geroch–Jang theorem does not depend on the detailed field equations for the metric, nor on the possible presence of additional gravitational degrees of freedom (e.g., the scalar field in the Brans–Dicke theory), nor on the number of dimensions of spacetime. Therefore, it is a promising starting point to test whether a generic metric theory of gravity satisfies or not the GWEP.

<sup>9</sup> The formulation of the theorem presented here is excerpted from Malament [18], where interesting remarks can be found about the niceties of the proposition and its hypotheses.

<sup>10</sup> That is, for all points  $P \in \mathcal{M}$ , and for all unit time-like vectors  $\xi^a$  at  $P$ , it is  $\Theta_{ab} \xi^a \xi^b \geq 0$ , and, if  $\Theta_{ab} \neq 0$ , then  $\Theta^a_b \xi^b$  is time-like.

<sup>11</sup> Of course, condition d) holds only if no forces other than gravity act on the body, which is precisely the case in free fall experiments.

Before we turn to the perturbative scheme, we must still discuss two issues potentially threatening the entire construction. First: the Geroch–Jang theorem implies that, if  $\Theta_{ab}$  satisfying conditions (a)–(d) exists, then the world line of a small body is a geodesic, hence the universality of free fall. Imagine, however, that we can find  $\Theta_{ab}$  satisfying (a)–(c) but not (d). What can be said about the resulting motion of the body?

Strictly speaking, we cannot claim that relaxing hypothesis (d) implies a nongeodesic path, because conditions (a)–(d) are only sufficient for the theorem to hold. Nevertheless, if one has good reasons to believe that  $\Theta_{ab}$  gives an appropriate description of the total energy-momentum content of the body — in the sense that its integral over a spacelike hypersurface represents the body four-momentum  $p_a$  — then one can adapt a derivation in Ref. [11], pp. 3382–3383, and argue that, if  $\bar{\nabla}^b \Theta_{ab} \neq 0$ , then an external force acts on the body and prevents its world line from being a geodesic. For this reason, in the rest of the paper we shall assume that the GWEP holds *if and only if* condition (4) is satisfied, in addition to (a)–(c), for a tensor  $\Theta_{ab}$  such that the difference  $\Theta_{ab} - T_{ab}$  represents the stress-energy-momentum associated with the gravitational field of the body under consideration.

The second issue deals with condition (c). If the body is spatially localised, then  $T_{ab}$  vanishes outside a world tube. But this is not the case for  $\Theta_{ab}$ , which contains also a contribution from the gravitational field. The dominant part of this contribution is the energy density of the static gravitational field associated with the body, which falls off as  $r^{-4}$ , where  $r$  is a suitable radial coordinate. Since this function has not a compact support,  $\Theta_{ab}$  fails to satisfy (c). To fix such issue, one has to notice that, in the proof of the Geroch–Jang theorem, condition (c) is used to guarantee that the integral over a spacelike hypersurface of a particular function proportional to  $\Theta_{ab}$  reduces to an integral performed only over the volume of the body. Calling  $M$  the mass of the body, and  $R$  the spatial size of the world tube associated with it, the contribution due to the external static gravitational field turns out to be proportional to  $GM^2/R$ . For a non-compact body, this is much smaller than the proper energy  $Mc^2$ , and even for an extreme object like a black hole it is sufficient to choose  $R$  much larger than the Schwarzschild radius — but still so small that the world tube can be well approximated by a world line — to ensure that no appreciable error arises when one truncates the range of integration at  $R$ . Therefore, in the following, we shall assume that a weaker form of the Geroch–Jang theorem holds, one which guarantees geodesic motion even under these conditions.<sup>12</sup>

<sup>12</sup> Even worse, there might be radiative contributions, whose energy density falls off as  $r^{-2}$ . As already pointed out in Sec. II A, we shall always restrict ourselves to experimental conditions where such terms can be neglected, so that the GWEP is at least satisfied by general relativity.

## B. Self-gravitation: The perturbative approach

We can now exhibit the actual tensor  $\Theta_{ab}$  required to test the validity of the GWEP for a given metric theory of gravity. To do so, we first observe that every such theory contains field equations that can always be cast in the form<sup>13</sup>

$$E_{ab} = T_{ab} . \quad (5)$$

Here, the stress-energy-momentum tensor  $T_{ab}$  can depend on the metric field  $g_{ab}$  and any matter field, but not on other possible gravitational degrees of freedom.<sup>14</sup> On the other hand, the tensor  $E_{ab}$  depends on all the gravitational variables and their derivatives. In the particular case of general relativity, for instance, it is

$$E_{ab} = \frac{1}{8\pi G} G_{ab} , \quad (6)$$

where  $G_{ab}$  is the Einstein tensor. If gravity is uniquely described by  $g_{ab}$ , we say that the theory is *purely metric*. If this is not the case, there are other gravitational field equations in addition to (5), associated with the other gravitational variables. Finally, there are the matter field equations.

Let us now suppose that a solution  $g_{ab}$  of Eq. (5) can be decomposed into the sum of two terms, namely a “background”  $\bar{g}_{ab}$ , and a “small perturbation”  $\epsilon \gamma_{ab}$ , where  $\epsilon$  is a bookkeeping parameter accounting for the correct orders of  $\gamma_{ab}$  in the expansions, so that

$$g_{ab} = \bar{g}_{ab} + \epsilon \gamma_{ab} \quad (7)$$

(see Appendix A for the corresponding expansion of several geometric quantities of interest). We can expand the tensors  $E_{ab}$  and  $T_{ab}$  accordingly, obtaining

$$E_{ab} = \bar{E}_{ab} + \epsilon \mathcal{E}_{ab} + E_{ab}^{(2+)} , \quad (8a)$$

$$T_{ab} = \bar{T}_{ab} + \epsilon \mathcal{T}_{ab} + T_{ab}^{(2+)} . \quad (8b)$$

<sup>13</sup> The very structure of our test for the GWEP makes it impossible to apply it also to a scalar theory of gravity. This leaves outside Nordström’s theory, which in turn is known to abide at least by the GNEP [19]. The reason is that, in such a theory, the dynamical scalar degree of freedom lives on a fixed, nondynamical Minkowskian background; this conflicts with our assumptions that only truly dynamical fields, viz. diffeomorphism-invariant structures, are allowed. As a consequence, one has that the equations of motion for test bodies cannot be derived from the field equations — as it happens, e.g., in general relativity [20], in view of its fully dynamical character — but must be instead postulated separately [4]. This issue, together with the impossibility of checking condition (d) in the Geroch–Jang theorem, prevents us from applying our method to a purely scalar theory of gravity.

<sup>14</sup> In the context of scalar-tensor theories (see Sec. IV B below), this is tantamount to choosing the so-called “Jordan frame”. In this way, non-self-gravitating particles follow the geodesics of the metric, and the WEP is satisfied — clearly, a prerequisite for the GWEP.

From this point onwards, symbols with an over bar are associated with quantities constructed out of background objects only. The script letters denote the parts which are linear in the perturbations, whereas the superscript “(2+)” generically denotes all the terms of order higher or equal than two in the expansions [10]. We assume furthermore that the background metric satisfies the zeroth-order field equations

$$\bar{E}_{ab} = \bar{T}_{ab} . \quad (9)$$

As it is normally done when testing the GWEP, we set  $\bar{T}_{ab} = 0$ ; that is, we ask the background solution to be a vacuum one; the crucial importance of such assumption will be fully explained at the end of Sec. III C. We can thus be sure that  $T_{ab}$  is made entirely of the small contribution due to the stress-energy-momentum tensor  $T_{ab}^{(p)}$  of the “particle” (i.e., the self-gravitating test body). Then, we can recast the field equations (5) in the equivalent form

$$\epsilon \mathcal{E}_{ab} = T_{ab}^{(p)} - E_{ab}^{(2+)} . \quad (10)$$

If now one can prove that

$$\bar{\nabla}^b \mathcal{E}_{ab} = 0 , \quad (11)$$

then the term  $E_{ab}^{(2+)}$  can be considered as a contribution due to the gravitational self-interaction of the body, and can be included into a new symmetric tensor

$$\Theta_{ab} = T_{ab}^{(p)} - E_{ab}^{(2+)} , \quad (12)$$

which is conserved with respect to the background in view of Eq. (11). This is all we need to invoke the Geroch–Jang theorem,<sup>15</sup> and hence deduce that the self-gravitating body is moving within a world tube which can be approximated with arbitrary precision by a geodesic of the background.

For all these reasons, we consider Eq. (11) to be a *necessary and sufficient* condition for the GWEP to hold.

### C. Variational formulation of the GWEP-test

The basic ingredients of our method — namely, equations (5), (8a), and (11) — might suggest at this stage that the actual check of the validity of the GWEP becomes only a matter of case-by-case calculations. For each metric theory of gravity, the different form of the tensors  $E_{ab}$  and  $\mathcal{E}_{ab}$  may in principle demand a separate

analysis. However, we shall see immediately that such a huge labour is not necessary, and can be effectively traded for a much more elegant and quicker inspection of the action of a given theory.

To this end, we notice first that, in a variational formulation of a theory of gravity, the tensors  $E_{ab}$  and  $T_{ab}$  in Eq. (5) are given by the functional derivatives, with respect to the inverse metric  $g^{ab}$ , of the gravitational action  $A_G$ , and of the matter action  $A_M$ , respectively:

$$E_{ab} = \frac{2}{\sqrt{-g}} \frac{\delta A_G}{\delta g^{ab}} ; \quad T_{ab} = -\frac{2}{\sqrt{-g}} \frac{\delta A_M}{\delta g^{ab}} . \quad (13)$$

Since the framework we are building is completely general,  $A_G$  is in principle a functional of the metric  $g_{ab}$ , and of any other gravitational degree of freedom  $\Phi_I$  ( $I$  represents a convenient set of indices). At the same time, the matter action  $A_M$  is a functional of  $g_{ab}$  and of the matter fields, but *not* of the  $\Phi_I$ , see the footnote commenting Eq. (5).

The field equations (5) then follow as soon as one requires that the total action  $A = A_G + A_M$  has a vanishing functional derivative with respect to  $g^{ab}$ . The matter field equations emerge from  $A_M$  alone when varied with respect to the matter fields. Similarly, additional gravitational field equations are derived varying  $A_G$  with respect to the  $\Phi_I$ . It is worth remarking here that the action  $A$  needs be equipped with all the apt boundary terms, necessary to derive the field equations in a consistent manner — that is, to guarantee that the functional derivatives of  $A$ , with respect to the various fields, exist. In the case of general relativity, for instance, one has to add the Gibbons–Hawking–York term [21, 22].

We can now decompose all the dynamical variables into a background part and a perturbation, the latter being associated with the parameter  $\epsilon$ , as in Sec. III B. The gravitational action can then be rewritten as

$$A_G = \bar{A}_G + \epsilon \mathcal{A}_G + A_G^{(2+)} , \quad (14)$$

where the zeroth-order term  $\bar{A}_G$  is a functional of the background fields only, whereas  $\mathcal{A}_G$  is a functional of the background and of the first-order perturbations. For convenience, we shall consider, as independent variables, the tensors  $\bar{g}^{ab}$  and  $\gamma^{ab}$ . This choice is entirely equivalent to the one in which the independent variables are  $\bar{g}_{ab}$  and  $\gamma_{ab}$ , but simplifies the treatment.

Varying Eq. (14) with respect to the background metric, and noticing that  $\gamma^{ab}$  is independent of  $\bar{g}^{ab}$ , one finds

$$\frac{\delta \bar{A}_G}{\delta \bar{g}^{ab}} + \epsilon \frac{\delta \mathcal{A}_G}{\delta \bar{g}^{ab}} + \frac{\delta A_G^{(2+)}}{\delta \bar{g}^{ab}} = \frac{\delta A_G}{\delta \bar{g}^{ab}} = \frac{\delta A_G}{\delta g^{cd}} \frac{\partial g^{cd}}{\partial \bar{g}^{ab}} = \frac{\delta A_G}{\delta g^{ab}} . \quad (15)$$

On the other hand, to the first order in  $\epsilon$ , it is

$$\frac{\delta A_G}{\delta g^{ab}} = \frac{\sqrt{-g}}{2} E_{ab} = \frac{\sqrt{-\bar{g}}}{2} \bar{E}_{ab} + \epsilon \frac{\sqrt{-\bar{g}}}{2} \left( \frac{\gamma}{2} \bar{E}_{ab} + \mathcal{E}_{ab} \right) , \quad (16)$$

<sup>15</sup> Strictly speaking, one should also check that  $\Theta_{ab}$  satisfies the other hypotheses in the theorem. However,  $T_{ab}$  satisfies them by assumption. Moreover, as discussed in Sec. III A, we suppose that the perturbation is sufficiently weak so that  $E_{ab}^{(2+)}$  does not alter condition (a), and that condition (c) can be relaxed without altering the conclusion.

where we have used Eqs. (8a) and (A3), and set  $\gamma := \bar{g}_{ab}\gamma^{ab}$ . Comparing Eqs. (15) and (16) we find

$$\bar{E}_{ab} = \frac{2}{\sqrt{-\bar{g}}} \frac{\delta \bar{A}_G}{\delta \bar{g}^{ab}}; \quad (17a)$$

$$\mathcal{E}_{ab} = \frac{2}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{g}^{ab}} - \frac{\gamma}{2} \bar{E}_{ab}. \quad (17b)$$

For the sake of definiteness, and without loss of generality, we suppose that all the non-metrical gravitational degrees of freedom reduce for the moment to a single scalar field  $\phi$  (the treatment can be straightforwardly extended to other geometric objects), and perform an analogous expansion

$$\phi = \bar{\phi} + \epsilon \chi + \phi^{(2+)} . \quad (18)$$

Then,  $\bar{A}_G[\bar{g}^{ab}, \bar{\phi}] = A_G[\bar{g}^{ab}, \bar{\phi}]$ , and since the matter action does not depend on  $\phi$ , the background scalar field  $\bar{\phi}$  must obey the equation

$$\frac{\delta \bar{A}_G}{\delta \bar{\phi}} = 0 . \quad (19)$$

Interestingly, it is also:

$$\frac{\delta \mathcal{A}_G}{\delta \gamma^{ab}} = -\frac{\sqrt{-\bar{g}}}{2} \bar{E}_{ab}; \quad (20a)$$

and

$$\frac{\delta \mathcal{A}_G}{\delta \chi} = \frac{\delta \bar{A}_G}{\delta \bar{\phi}} . \quad (20b)$$

These relations appear obvious, as soon as one notices that  $\epsilon \mathcal{A}_G$  is just the first-order variation of  $A_G$  corresponding to the variations  $\delta g^{ab} = -\epsilon \gamma^{ab}$  and  $\delta \phi = \epsilon \chi$  of the fields. It is instructive, however, to see a formal proof. An explicit calculation of  $\mathcal{A}_G$  gives

$$\begin{aligned} \mathcal{A}_G[\bar{g}^{ab}, \gamma^{ab}, \bar{\phi}, \chi] &= \left( \frac{dA_G[g^{ab}, \phi]}{d\epsilon} \right)_{\epsilon=0} = \int_U d^m x \left[ \left( \frac{\delta A_G}{\delta g^{ab}} \right)_{\epsilon=0} \left( \frac{dg^{ab}}{d\epsilon} \right)_{\epsilon=0} + \left( \frac{\delta A_G}{\delta \phi} \right)_{\epsilon=0} \left( \frac{d\phi}{d\epsilon} \right)_{\epsilon=0} \right] \\ &= \int_U d^m x \left( -\frac{\delta \bar{A}_G}{\delta \bar{g}^{ab}} \gamma^{ab} + \frac{\delta \bar{A}_G}{\delta \bar{\phi}} \chi \right) = \int_{\mathcal{U}} d\bar{\Omega} \left( -\frac{1}{2} \bar{E}_{ab} \gamma^{ab} + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \bar{A}_G}{\delta \bar{\phi}} \chi \right), \end{aligned} \quad (21)$$

where  $\mathcal{U}$  and  $U$  denote, respectively, the domain of integration and its coordinate representation in  $\mathbb{R}^m$  (for simplicity, we assume that  $\mathcal{U}$  can be covered by a single chart), and  $d\bar{\Omega} = \sqrt{-\bar{g}} d^m x$  is the volume element with respect to the background metric. Equations (20a) and (20b) then follow immediately.

It is important to realise that this expression for  $\mathcal{A}_G$  is correct provided that the functional derivatives of  $A_G$  with respect to the gravitational fields  $g^{ab}$  and  $\phi$  exist.

As is well-known, this is the case provided that suitable conditions on the fields themselves and their derivatives are imposed on the boundary  $\partial\mathcal{U}$ . Of course, this basically amounts to requiring that the variational principle be well-formulated for the theory under scrutiny.

Consider now a diffeomorphism, generated by the vector field  $\xi^a$ . Order by order in  $\epsilon$ , diffeomorphism invariance of the gravitational theory implies  $\delta \bar{A}_G = 0$ ,  $\delta \mathcal{A}_G = 0$ , and so forth. In particular, the second of these conditions becomes, with standard manipulations,

$$\begin{aligned} 0 &= \int_U d^m x \left( \frac{\delta \mathcal{A}_G}{\delta \bar{g}^{ab}} \delta \bar{g}^{ab} + \frac{\delta \mathcal{A}_G}{\delta \gamma^{ab}} \delta \gamma^{ab} + \frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} \delta \bar{\phi} + \frac{\delta \mathcal{A}_G}{\delta \chi} \delta \chi \right) \\ &= \int_{\mathcal{U}} d\bar{\Omega} \left( -\frac{2}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{g}^{ab}} \bar{\nabla}^b \xi^a + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \gamma^{ab}} \delta \gamma^{ab} + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} \xi^a \bar{\nabla}_a \bar{\phi} \right), \end{aligned} \quad (22)$$

where  $\delta \gamma^{ab} = \xi^c \bar{\nabla}_c \gamma^{ab} - \gamma^{cb} \bar{\nabla}_c \xi^a - \gamma^{ac} \bar{\nabla}_c \xi^b$ , and we have dropped a term using Eq. (20b) together with the field equation (19) for the background scalar field  $\bar{\phi}$ . Integrating by parts, and eliminating a boundary term choosing the arbitrary vector field  $\xi^a$  such that it vanishes on  $\partial\mathcal{U}$ , we find

$$\begin{aligned} 0 &= \int_{\mathcal{U}} d\bar{\Omega} \left[ \left( \bar{\nabla}^b \left( \frac{2}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{g}^{ab}} \right) + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} \bar{\nabla}_a \bar{\phi} \right) \xi^a + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \gamma^{ab}} \delta \gamma^{ab} \right] \\ &= \int_{\mathcal{U}} d\bar{\Omega} \left( \bar{\nabla}^b \mathcal{E}_{ab} + \frac{1}{\sqrt{-\bar{g}}} \frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} \bar{\nabla}_a \bar{\phi} - \frac{1}{2} \bar{E}_{bc} \bar{\nabla}_a \gamma^{bc} - \bar{\nabla}_c (\bar{E}_{ab} \gamma^{cb}) + \frac{1}{2} \bar{\nabla}^b (\gamma \bar{E}_{ab}) \right) \xi^a, \end{aligned} \quad (23)$$



where we have used Eqs. (17b) and (20a). By the arbitrariness of  $\xi^a$  we find, finally,

$$\bar{\nabla}^b \mathcal{E}_{ab} = -\frac{1}{\sqrt{-g}} \frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} \bar{\nabla}_a \bar{\phi} + \frac{1}{2} \bar{E}_{bc} \bar{\nabla}_a \gamma^{bc} + \bar{\nabla}_b (\bar{E}_{ac} \gamma^{bc}) - \frac{1}{2} \bar{E}_{ab} \bar{\nabla}^b \gamma. \quad (24)$$

In the last term above, we have used the property  $\bar{\nabla}^b \bar{E}_{ab} = 0$ , which follows by diffeomorphism invariance of  $\bar{\mathcal{A}}_G$  together with the field equation (19) for  $\bar{\phi}$ .

Since  $\bar{\phi}$  and  $\gamma_{ab}$  are independent, equation (24) implies that, in order for Eq. (11) to hold, the following two conditions must be satisfied simultaneously:<sup>16</sup>

$$\frac{\delta \mathcal{A}_G}{\delta \bar{\phi}} = 0; \quad (25a)$$

$$\bar{E}_{ab} = 0. \quad (25b)$$

These are rather different in character, so we comment on them separately.

Condition (25a) is “structural”, i.e., it concerns the type of theory considered. One can realise that, even assuming the validity of the field equations, the quantity  $\delta \mathcal{A}_G / \delta \bar{\phi}$  vanishes only if the gravitational action does not depend on the variable  $\phi$ . Dropping the initial restriction of a single additional scalar field, this is tantamount to requiring that one is dealing with a purely metric theory of gravity. Of course, a constant background scalar field  $\bar{\phi}$  would still make the first term on the right-hand side of Eq. (24) vanish. This circumstance, however, can occur only for extremely special background metrics, hence it is irrelevant for the general validity of the GWEP within a given theory.

On the other hand, condition (25b) is “environmental”. Using the field equations (9), it implies that condition (11) can be satisfied only in an empty background (that is, for  $\bar{T}_{ab} = 0$ ). This assumption is always made in the context of tests of the GWEP or of the SEP, and it is interesting to find that it emerges from our formalism as a necessary condition for the GWEP. Let us then clarify its role.

Suppose that  $\bar{T}_{ab} \neq 0$  for some otherwise unspecified background. A self-gravitating body produces, by definition, a non-negligible gravitational field in its surroundings, and this exerts, in general, a gravitational force on the background matter. By the action-reaction principle, there is also a corresponding force on the body itself, which makes its world line deviate from a geodesic of  $\bar{g}_{ab}$ . A Newtonian estimate for the potential energy

density associated with these forces is  $\bar{\rho} \Phi$ , where  $\Phi$  is the gravitational potential due to the body, and  $\bar{\rho}$  is the mass density of the background. The relativistic tensorial generalisation of  $\bar{\rho} \Phi$  is  $\bar{T}_{ac} \gamma^{cb}$ , and considering the field equations (9) for the background, this explains the appearance of the terms in Eq. (24). The derivatives are, of course, due to the fact that the quantity evaluated in Eq. (24) is related to the force density, rather than to the potential energy density.

To sum up, an empty background is a necessary condition for the GWEP to hold, because in the presence of matter the test body couples to it via its own gravitational field, and this produces, in general, a force that violates the universality of free fall. Thus,  $\bar{T}_{ab} = 0$  is one of the basic conditions (together with the assumptions that the body is spatially localised, and that it affects the environment negligibly) that must be postulated so that the very issue of whether a theory does, or does not, satisfy the GWEP makes sense at all.

This conclusion receives additional, independent support if one notices that the presence of matter at the background level would prevent the identification  $T_{ab} = T_{ab}^{(p)}$ , which was crucial for writing Eq. (10). This would imply that one is no longer able to confine the generalised stress-energy-momentum tensor  $\Theta_{ab}$  given by Eq. (12) to the interior of a world tube, as it will depend not only on the particle, but in general also on the distribution of background matter. This fact in turn violates one of the assumptions required for the Geroch–Jang theorem to hold, thus shaking the very foundations of our framework.

#### IV. CASE STUDIES

We now investigate explicitly whether condition (11) for the GWEP holds in some relevant classes of metric theories. This analysis complements the general results obtained in Sec. III C, culminating in Eq. (24), and allows for a further discussion of some niceties. We first consider two cases — general relativity and the Brans–Dicke theory — in which the validity of the GWEP has been already confirmed or disproved by means of other theoretical and experimental arguments. In this sense, such theories are used here as preliminary sanity checks.

After that, we move on to three other cases where the status of the principle is less certain, or even completely unknown. Specifically, we examine general scalar-tensor theories, then higher-curvature and higher-order  $f(R)$  theories, and finally the higher-curvature but not higher-order Lanczos–Lovelock theories. The flexibility

<sup>16</sup> These conditions are unavoidable. In particular, one might try to escape condition (25b) making the last three terms in Eq. (24) vanish, exploiting the gauge freedom  $\gamma_{ab} \rightarrow \gamma_{ab} + \bar{\nabla}_a \zeta_b + \bar{\nabla}_b \zeta_a$ , where  $\zeta_a$  is arbitrary. Of course, this will not succeed, because the independent terms to be killed are more than the four arbitrary functions provided by  $\zeta_a$ . Moreover, such a transformation would also affect  $\mathcal{E}_{ab}$ , thus making other terms appear on the left-hand side of Eq. (24).

and generality of condition (11), together with the variational formalism, will show all its power precisely in such more exotic scenarios, where other theoretical arguments turn out to be inadequate or even completely inapplicable.<sup>17</sup>

The simple “yes/no” answer one can extract from a test of the GWEP, however, is not the only relevant bit of information one can get. By this analysis, we can also try to better understand in which precise sense the fact that a given theory satisfies or not the GWEP has to be linked with the existence of additional gravitational degrees of freedom, as it happens in the Brans–Dicke theory, or can be also attributed to the presence of higher-curvature terms in the action, as in  $f(R)$  and the Lanczos–Lovelock theories.

### A. General relativity

Being the test-bench upon which the SEP (and, thus, the GWEP) has been tailored, it is evident that Einstein’s tensor theory of gravity in four spacetime dimensions abides by such principle — of course, with all the caveats discussed in Secs. II A and III A. However, given the features of our protocol, the restriction to four-dimensional manifolds can be set aside, and we can test the validity of the GWEP on any  $m$ -dimensional version of general relativity (with  $m \geq 4$ ). Hence, in the present section we work directly in the most general case.

The Einstein–Hilbert action (with the addition of the Gibbons–Hawking–York boundary term [21–23]) underlying the gravitational sector of the theory is

$$A_{\text{GR}} = \frac{1}{16\pi G} \left( \int_{\mathcal{U}} d\Omega (R - 2\Lambda) + 2 \oint_{\partial\mathcal{U}} d\Sigma K \right). \quad (26)$$

Here,  $R$  is the scalar curvature,  $\Lambda$  is a cosmological constant<sup>18</sup> (considered as a fundamental constant of Nature, but with its value left unspecified),  $d\Sigma = d^{m-1}x \sqrt{|h|}$  is the volume element on the  $(m-1)$ -dimensional boundary  $\partial\mathcal{U}$  ( $h$  is the determinant of the induced metric), and  $K$  is the trace of the extrinsic curvature. We notice that the presence of the boundary term is strictly necessary to make the variational problem well posed, as it permits for an identical cancellation of another boundary term

containing  $\nabla_c \delta g_{ab}$  that appears when the metric is varied by  $\delta g_{ab}$ . Then, to get Einstein’s field equations, all one has to set is  $\delta g_{ab} = 0$  at the boundary [6, 23]. This leads to the following expression for the tensor  $E_{ab}$ :

$$E_{ab} = \frac{1}{8\pi G} (G_{ab} + \Lambda g_{ab}). \quad (27)$$

From the results of Sec. III C, it is obvious that this theory, being purely metric, passes our test of the GWEP. However, the validity of condition (11) can also be checked independently, via a full calculation of  $\mathcal{E}_{ab}$  and its covariant divergence with respect to the background metric. Since, at first order,

$$\mathcal{E}_{ab} = \frac{1}{8\pi G} (G_{ab} + \Lambda \gamma_{ab}), \quad (28)$$

where  $\epsilon G_{ab}$  is the first-order term in the expansion of the Einstein tensor (see Eq. (A11) for the full-length expression), condition (11) reduces simply to

$$\bar{\nabla}^b \mathcal{G}_{ab} = -\Lambda \bar{\nabla}^b \gamma_{ab}. \quad (29)$$

A somewhat lengthy, but straightforward calculation (see Appendix B) shows that the quantity  $\bar{\nabla}^b \mathcal{G}_{ab}$  is given by Eq. (B5). However, as stated above, we work on a background which is a solution of the vacuum field equations, so  $\bar{R}_{ab} = \Lambda \bar{g}_{ab}$ . Plugging this into the general expression (B5), and using the metric-compatibility condition  $\bar{\nabla}_c \bar{g}_{ab} = 0$ , we find that Eq. (29) is indeed satisfied.

This assures that, in Einstein’s theory with an arbitrary cosmological constant (hence, in particular, also for  $\Lambda = 0$ ), the GWEP holds true, and the world lines of bodies with negligible size but non-negligible self-gravitation in an empty background are geodesics.

### B. Scalar-tensor theories

Let us now apply our method to a prototypical example of an extended theory of gravity; that is, one going beyond the framework of general relativity. In this particular case, it is already known that the theory does not obey the GNEP. We consider the Brans–Dicke theory [24], which corresponds to the gravitational action

$$A_{\text{BD}} = \frac{1}{16\pi} \int_{\mathcal{U}} d\Omega \left( \phi R - \frac{\omega}{\phi} g^{ab} \nabla_a \phi \nabla_b \phi \right) + \text{b.t.}, \quad (30)$$

where  $\omega$  is a nonzero constant and, here and in the following, “b.t.” stands for an appropriate boundary term. The main feature of this theory is the presence of an additional gravitational degree of freedom, in the form of a single scalar field  $\phi$ . A comparison with the action (26) for general relativity shows that we have simply replaced the constant  $1/G$  with a function  $\phi$ , and then given such field a dynamical behaviour via the additional kinetic

<sup>17</sup> The standard PPN formalism, for instance, is tailored on a four-dimensional spacetime, and becomes of no practical utility (unless suitably extended) in higher dimensions, which is what happens precisely in the context of Lanczos–Lovelock theories of gravity.

<sup>18</sup> The case of a  $\Lambda$  term has not been investigated thoroughly using the PPN formalism, mostly because of its extremely tiny observational value (of order  $10^{-122}$  in Planck units), which is usually considered small enough to be neglected in the context of solar system experiments. Hence, it remains unclear whether the addition of such a term might cause the traditional PPN implementation of the SEP/GNEP to fail, albeit in practice one would not expect any significant violation to arise.

term. Thus, we have formally promoted Newton's constant  $G$  to a function on spacetime. As in the case of Einstein's theory, the boundary term makes the variational formulation well posed.

By varying  $A_{\text{BD}}$  above with respect to  $g^{ab}$ , one gets field equations of the form (5), with

$$E_{ab} = \frac{1}{8\pi} \left( \phi G_{ab} - \frac{\omega}{\phi} \left( \nabla_a \phi \nabla_b \phi - \frac{1}{2} g_{ab} \nabla^c \phi \nabla_c \phi \right) - \nabla_a \nabla_b \phi + g_{ab} \square \phi \right), \quad (31)$$

where  $\square := g^{ab} \nabla_a \nabla_b$  is the d'Alembertian operator. In addition, there is another scalar equation accounting for the dynamics of the long-range field  $\phi$ , namely

$$R - \frac{\omega}{\phi^2} \nabla^a \phi \nabla_a \phi + \frac{2\omega}{\phi} \square \phi = 0. \quad (32a)$$

This, together with Eq. (31), yields

$$\square \phi = \frac{8\pi}{3 + 2\omega} T, \quad (32b)$$

where  $T = g^{ab} T_{ab}$ . Equations (32a) and (32b) can be used interchangeably.

Since this theory contains the scalar field  $\phi$  in its gravitational sector, it is not purely metric and violates condition (25a). Hence, condition (11) is also violated, and the GWEP does not hold. It is instructive, however, to consider this result from another point of view. Let us

pick a background vacuum solution, so  $\bar{T}_{ab} = 0$  and condition (25b) is satisfied. The zeroth-order field equation  $\bar{E}_{ab} = 0$  can be rearranged as  $\bar{G}_{ab} = \bar{T}_{ab}^{(\phi)}$ , where

$$\bar{T}_{ab}^{(\phi)} = \frac{\omega}{\bar{\phi}^2} \left( \bar{\nabla}_a \bar{\phi} \bar{\nabla}_b \bar{\phi} - \frac{1}{2} \bar{g}_{ab} \bar{\nabla}^c \bar{\phi} \bar{\nabla}_c \bar{\phi} \right) + \frac{\bar{\nabla}_a \bar{\nabla}_b \bar{\phi}}{\bar{\phi}} - \bar{g}_{ab} \frac{\bar{\square} \bar{\phi}}{\bar{\phi}}, \quad (33)$$

with the indices raised with, and the d'Alembertian operator built out of, the background metric alone. Thus,  $\bar{T}_{ab}^{(\phi)}$  can be thought of as a “matter” stress-energy-momentum tensor associated with the dynamical long-range field  $\phi$  at zeroth order. By such redefinition, we see immediately that we fall back into something which is, formally, just general relativity with a particular non-empty background — hence the failure of the GWEP.

We can summarise the results as follows: for Brans–Dicke gravity, the existence of an additional gravitational degree of freedom besides the metric, i.e., the non-purely metric character of the theory, leads to the failure of the GWEP because of the non-vanishing first term in Eq. (24). Alternatively, upon rearranging the field equations so as to have some sort of “pure general relativity plus exotic scalar matter”, the other terms in Eq. (24) spoil the equivalence, and once again certify the failure of the geodesic motion for self-gravitating test bodies.

A straightforward calculation of  $\bar{\nabla}^b \mathcal{E}_{ab}$ , as done in the case of general relativity, would be very lengthy and cumbersome. Just to convey an idea of the sort of objects involved, and of the power of the general approach developed in Sec. III C, we exhibit the explicit form of the first-order tensor  $\mathcal{E}_{ab}$ , which reads

$$\begin{aligned} \mathcal{E}_{ab} = & \frac{1}{8\pi} \left( \bar{\phi} \mathcal{G}_{ab} + \chi \bar{G}_{ab} - \frac{\omega}{\bar{\phi}} (\bar{\nabla}_a \bar{\phi} \bar{\nabla}_b \chi + \bar{\nabla}_a \chi \bar{\nabla}_b \bar{\phi}) + \frac{\omega \chi}{\bar{\phi}^2} \bar{\nabla}_a \bar{\phi} \bar{\nabla}_b \bar{\phi} - \frac{\omega}{2\bar{\phi}} \bar{g}_{ab} \gamma^{cd} \bar{\nabla}_c \bar{\phi} \bar{\nabla}_d \bar{\phi} \right. \\ & + \frac{\omega}{2\bar{\phi}} \gamma_{ab} \bar{g}^{cd} \bar{\nabla}_c \bar{\phi} \bar{\nabla}_d \bar{\phi} + \frac{\omega}{\bar{\phi}} \bar{g}_{ab} \bar{g}^{cd} \bar{\nabla}_c \bar{\phi} \bar{\nabla}_d \chi - \frac{\omega \chi}{2\bar{\phi}^2} \bar{g}_{ab} \bar{g}^{cd} \bar{\nabla}_c \bar{\phi} \bar{\nabla}_d \bar{\phi} + \Xi^c_{ab} \bar{\nabla}_c \bar{\phi} \\ & \left. - \bar{g}_{ab} \gamma^{cd} \bar{\nabla}_c \bar{\nabla}_d \bar{\phi} - \bar{g}_{ab} \bar{g}^{cd} \Xi^e_{cd} \bar{\nabla}_e \bar{\phi} + \bar{g}_{ab} \bar{g}^{cd} \bar{\nabla}_c \bar{\nabla}_d \chi + \gamma_{ab} \bar{g}^{cd} \bar{\nabla}_c \bar{\nabla}_d \bar{\phi} - \bar{\nabla}_a \bar{\nabla}_b \chi \right), \end{aligned} \quad (34)$$

where  $\Xi^a_{bc}$  is given by Eq. (A5),  $\mathcal{G}_{ab}$  is given by Eq. (B1), and  $\bar{G}_{ab}$  is related to  $\bar{\phi}$  and its derivatives through the background field equation  $\bar{G}_{ab} = \bar{T}_{ab}^{(\phi)}$ , with  $\bar{T}_{ab}^{(\phi)}$  given by Eq. (33).

Upon generalising the action (30) to the one of an arbitrary scalar-tensor theory, namely

$$A_{\text{ST}} = \frac{1}{16\pi} \int_{\mathcal{M}} d\Omega \left( \phi R - \frac{\omega(\phi)}{\phi} \nabla^a \phi \nabla_a \phi + V(\phi) \right) + \text{b.t.}, \quad (35)$$

the conclusion about the failure of the GWEP clearly does not change, as the theory is still not purely metric. By the same token, one might say that, upon rewrit-

ing the field equations as “general relativity with scalar matter”, the generic zeroth-order tensor  $\bar{T}_{ab}^{(\phi)}$  would still prevent the validation of the GWEP.

Noticeably, the same logic can be applied to the much wider class of multi-scalar-tensor theories [24]. Either the appearance of more than one additional scalar degree of freedom in (24) (where the first term ought to be traded for a sum over the number of scalar fields available), or the presence of a cumulative nonzero background stress-energy tensor for the scalar fields — also in matter vacuum — will lead to the failure of condition (11).

### C. Higher-order theories and their subtleties

The last paragraphs above allow us to discuss yet another interesting case, viz., a vast sub-class of the so-called higher-order theories of gravity, for which the true status of the GWEP is still somewhat debated [1]. As we shall see, while the core of the method used so far remains unchanged, there are some crucial niceties worth pointing out, to shine a light on some misconceptions afflicting the zoology of higher-order actions.

We begin with the basics. In the simplest cases of higher-curvature theories, the action is built substituting the scalar curvature  $R$  in Eq. (26), with a more general arbitrary function  $f(R)$ , supposed to be analytic in its argument. The resulting field equations — if existing at all; see below — generally exhibit derivatives of order higher than 2, which can be sometimes reformulated in terms of excitations of additional modes of the graviton propagator and other physically measurable effects [25]. The next step is to identify even further scalar combinations of  $R_{abc}{}^d$ ,  $R_{ab}$ ,  $R$ , and functions thereof, exhausting the landscape of higher-curvature theories.

At a first glance, it might appear that the only gravitational degrees of freedom are those encoded in the metric, so that e.g., a generic  $f(R)$  theory is purely metric, like general relativity. Hence, the entire class of theories should satisfy the GWEP. The same argument might then be repeated identically for more general cases.

This conclusion is actually wrong. The reason, however, is partially concealed by the way the gravitational actions are built, and only a deeper inspection allows for a full understanding of the pitfall. The key element is the role played by the boundary terms in the action.

Let us start with the exemplary case of an  $f(R)$  theory. The action can be written as

$$A_{f(R)} = \frac{1}{16\pi} \int_{\mathcal{U}} d\Omega f(R) + \text{b.t.}, \quad (36)$$

where the boundary term basically mirrors that of the Einstein–Hilbert action [26]. Upon performing the variation, the resulting field equations read

$$f'(R) R_{ab} - \frac{1}{2} f(R) g_{ab} - \nabla_a \nabla_b f'(R) + g_{ab} \square f'(R) = 8\pi T_{ab}, \quad (37)$$

where  $f'(R) := df(R)/dR$ . To get this result, however, one is forced to impose the condition  $\delta R = 0$  on the boundary [26], which amounts to fixing the value of a combination of the metric and its first and second derivatives. This seems too strong a requirement, even for a fourth-order theory as this one. It thus seems that the functional derivative with respect to  $g^{ab}$  of the action (36) is not defined, and that the field equations (37) cannot be derived by a variational principle, after all.

The way out of this difficulty is to recognise that, in fact, there is another degree of freedom, associated with a scalar field  $\phi$ , which is kinematically independent of  $R$

but dynamically related to it in an algebraic manner, so that  $\delta R = 0$  is not a condition on the second derivatives of the metric, but rather on  $\phi$ . The action (36) can be equivalently rewritten as that of a particular scalar-tensor theory where one of the field equations is just  $\phi := f'(R)$  [27]. Only in this new formulation can the boundary term be introduced and all problems removed.<sup>19</sup> The variational principle becomes then well posed and our treatment in Sec. III C applies. Thus, the seemingly purely metric  $f(R)$  theories of gravity unveil themselves as generalised scalar-tensor theories in disguise [25], and do not pass our test for the GWEP.

This scheme can be generalised. All known cases of theories with higher-order field equations require some additional constraints on the boundary (the “higher momenta”) to get the field equations from the action. Such extra conditions go beyond the standard ones expected from a purely metric framework, and could be avoided only introducing boundary integrals analogous to the Gibbons–Hawking–York counter-term of general relativity. No such boundary terms, however, have been found yet for such theories. At the same time, theories of the kind  $f(R, \square R, \dots)$  — which indeed produce higher-order field equations — have been suitably reformulated into equivalent multi-scalar-tensor theories with second-order field equations only [28–30]. The variational formulation of these equivalent theories is well posed, but the presence of additional degrees of freedom implies that the GWEP is violated.

These results suggest that a theory with higher curvatures in the action, but only second-order field equations, may provide a different answer, and enlarge the category of GWEP-validating frameworks. The question might seem bizarre, as  $A_{\text{GR}}$  given by Eq. (26) is the only non-trivial action one can write in four dimensions, which gives second-order field equations (while being also linear in the curvature). This issue, however, is not so trivial in higher dimensions. It is known since 1971 that there is a whole class of higher-curvature theories, described by the so-called Lanczos–Lovelock actions [31], leading to at most second-order field equations. For example, in five dimensions, besides the 5D-version of general relativity, one has also Gauß–Bonnet gravity [32], which has an action characterised by terms quadratic in the curvature. This theory (and all its higher-dimensional companions) is therefore an optimal test bench for discriminating another aspect that might be crucial in the realisation or failure of the GWEP.

<sup>19</sup> Already at the level of elementary, one-dimensional mechanical systems, it is possible to convince oneself that the lack of a proper boundary term, or the need to impose too many conditions at the boundary, can easily lead to any sort of inconsistencies, loop arguments, and ill-posedness of the equations of motion [23].

### D. Gauß–Bonnet and Lanczos–Lovelock gravity

The general Lanczos–Lovelock gravitational action in an  $m$ -dimensional spacetime has the compact form [32]

$$A_{\text{LL}} = \frac{1}{16\pi G} \int_{\mathcal{U}} d\Omega Q_{abcd} R^{abcd} + \text{b.t.} \quad (38)$$

The tensor  $Q_{abcd}$  retains all the algebraic symmetries of the Riemann tensor, and has vanishing divergence,  $\nabla_d Q_{abc}{}^d = 0$ . Such object is designed so as to provide, for each value of  $m$ , all the combinations giving rise to second-order field equations, but not higher derivative terms. This feature, together with variational arguments, makes it possible to recognise the tensor  $Q_{abcd}$  as a polynomial in the curvature tensor with constant coefficients, whose degree depends on the dimension of spacetime. The zeroth-order term of such polynomial is simply the combination  $g_{a[c} g_{d]b}$ . When contracted with the Riemann curvature tensor, this gives the scalar curvature  $R$ , thus leading to the Einstein–Hilbert action (26).

The search for the boundary terms is in this case particularly labourious. Yet, such terms exist [33–36], and make the variational problem well posed for all Lanczos theories. As a consequence, and this is of course relevant for our purposes, Lovelock gravity is a purely metric theory.<sup>20</sup>

The lowest nontrivial element of the Lanczos–Lovelock class is the Gauß–Bonnet theory — emerging from  $m = 5$  onwards — for which it is

$$Q_{abcd} = \alpha (R_{abcd} + G_{bc}G_{ad} - G_{ac}G_{bd} + R_{ad}G_{bc} - R_{bd}G_{ac}), \quad (39)$$

where  $\alpha$  is a constant with dimensions (length)<sup>2</sup>. The corresponding Lagrangian density is

$$Q_{abcd}R^{abcd} = \alpha (R^2 - 4R^{ab}R_{ab} + R^{abcd}R_{abcd}). \quad (40)$$

When one also adds the zeroth-order term  $R$ , the variation of the action with respect to the metric gives

$$E_{ab} = \frac{1}{8\pi G} (G_{ab} + \alpha Y_{ab}), \quad (41)$$

where the tensor  $Y_{ab}$  is

$$Y_{ab} := 2RR_{ab} - 4R_{ac}R_b{}^c - 4R^{cd}R_{acbd} + 2R_a{}^{cde}R_{bcde} - \frac{1}{2}g_{ab}(R^2 - 4R^{cd}R_{cd} + R^{cdef}R_{cdef}). \quad (42)$$

<sup>20</sup> There has been a recent attempt [37] to describe Lanczos–Lovelock actions in terms of pure  $m$ -dimensional general relativity, plus a convenient number of 3-form fields and their covariant derivatives. Yet, an accurate inspection of Ref. [37] — particularly of Eq. (21) therein — shows that the new fields supposedly encoding the additional gravitational degrees of freedom have identically vanishing variation, which makes them just auxiliary variables, rather than truly dynamical ones.

An analogous recipe permits to build all the subsequent Lanczos–Lovelock terms compatible with the given space-time dimension. The higher-curvature corrections to the Einstein–Hilbert action are all tailored so that no field equations of order higher than 2 can emerge.

Since, as noticed, any Lanczos–Lovelock theory is purely metric, condition (25a) is satisfied, hence  $\bar{\nabla}^b \mathcal{E}_{ab} = 0$  in vacuum. Therefore, the GWEP is valid within the entire class of such theories. This last result has been advanced in Ref. [38] for the Gauß–Bonnet theory, as a consequence of a completely different point of view — namely, that of Noether-like conserved currents.

There are some interesting consequences. First, we have now another class of theories fulfilling the requirements to obey the GWEP. Such theories complement general relativity (actually, its  $m$ -dimensional version, for each  $m \geq 4$ ), which was already known to abide by the principle [19, 38].

Furthermore, and this is perhaps more interesting, for  $m$  strictly greater than 4, there are several competing purely metric theories of gravity, all providing an almost-geodesic motion for self-gravitating test bodies. Hence, there might be other criteria than the GWEP, playing a role in further selecting a subset within this class. Presumably, they will have to deal with the specific form of gravitational self-coupling.

In this sense, once again Lanczos–Lovelock theories prove to be “special” within the landscape of extended theories of gravitation, and a “natural” generalisation of Einstein’s seminal model [32, 36]. At the same time, the difference in the actual construction of the gravitational action, and the resulting different phenomenology [39, 40], seem to indicate that other criteria may possibly be introduced, to build an even finer taxonomic sieve.

### V. CONCLUDING REMARKS

In this work, we have shown how it is possible to find novel and useful bits of information in the seemingly outdated principle of equivalence, and we have built a formal and operative protocol to single out purely metric theories of gravitation starting from heuristic concepts and elementary considerations about the behaviour of self-gravitating test bodies.

Notwithstanding the inevitably perturbation-based scheme, and the number of hypotheses and restrictions necessary for the formulation to be meaningful, we deem to have finally given the gravitational weak equivalence principle a truly constructive role, similar to that of other equivalence principles such as the WEP and EEP, in the selection of gravitation theories.

The technique proposed overcomes at once the difficulties and limitations intrinsic to the PPN formalism (and similar methods developed for solar system experiments or for astronomical conditions alike), relying on a much more immediate inspection of the first-order action

for a given theory of gravity, provided that such action is equipped with boundary terms that make the variational principle well posed. It can be adapted seamlessly to spacetimes of arbitrary dimension, thus permitting a quick access to all those theories for which the PPN formalism and similar procedures have not yet been developed.

Our conclusions fully agree with the common interpretation of the strong equivalence principle (or rather, with a restricted version of it), in the sense that only within purely metric theories of gravity does universality of free fall extend to self-gravitating bodies [1], and only if the background environment is freed from matter content other than the body under consideration [10]. However, we provide a formally sound understanding of the roots of these generally accepted notions. Our method confirms or disproves the GWEP in any case where other independent techniques provide an answer. This has been shown working out explicitly the examples of general relativity (with or without a cosmological constant term), the Brans–Dicke theory, and general (multi-)scalar-tensor theories. But we have also presented new results, concerning higher-order and Lanczos–Lovelock theories.

Since the GWEP is ultimately satisfied only by purely metric theories, testing it becomes a matter of acknowledging the existence of extra gravitational degrees of freedom. The latter are not always evident, but can be hidden in the structure of the action (as it happens for the higher-order theories discussed in Sec. IV C), eventually cropping up through the impossibility of adding proper boundary terms that make the functional derivative with respect to the metric well defined. On the other hand, a failure of the GWEP may be considered a further hint that such hidden variables are present, and that only a rethinking of the theories in terms of more degrees of freedom will lead to their variational formulation becoming well posed. Thus, not only is the GWEP singling out purely metric theories; it can also become an independent tool to unveil the deep structure of gravity theories, which is sometimes concealed by a particular choice of variables.

An issue that remains untouched is the relation between our findings and the recent proposal in Refs. [14, 15]. It is difficult, at this stage, to say whether some form of general agreement could be reached, because the results of the two approaches can be compared only in the cases of general relativity and scalar-tensor theories (where they agree, of course). Indeed, the extension of condition (2) to more general scenarios is not provided in the original references, and might well require adjustments when passing, e.g., to the context of Lanczos–Lovelock gravity.<sup>21</sup>

As an aside, since in higher-dimensional spacetimes it has been found that not only general relativity, but also the whole class of higher-curvature but not higher-derivative Lanczos–Lovelock theories comply with the GWEP, there is potential room for improvement in the selection rule, perchance integrating our test with some restrictions on the kind of self-interaction of the gravitational degrees of freedom. In this respect, it is worth noticing that the  $m$ -dimensional version of Einstein’s theory contains the minimal self-coupling conceivable in a non-linear context, whereas the complexity increases with the order of the Lanczos–Lovelock polynomials. It is far from clear, though, whether and how this might be implemented thoroughly as an additional selection criterion.

Our work has attempted at repeating, on a much smaller scale, the historical process through which a set of heuristic statements was turned into a formal synthesis able to motivate precision-test experiments. In doing so, we have further refined our understanding of some of the relations among the various extended theories of gravity, and partially satiated a “taxonomic hunger” which is germane to the community of researchers [27]. Far from being a dry exercise in abstract systematics, this urge to detail the family tree of gravitational theories is motivated by both puzzling observational evidence of an unexplained behaviour of our universe [41], and daring theoretical proposals (e.g., the AdS/CFT duality). And it is somewhat comforting to see that the most fundamental, although perhaps slightly naïve, intuitions of our predecessors can provide, still nowadays, robust guidelines for a better understanding of gravity.

## ACKNOWLEDGMENTS

E.D.C. gratefully acknowledges: Vincenzo Vitagliano, Goffredo Chirco, Iwona Mochol, and Arletta Nowodworska, for their inspiring support; Eric Poisson, Ian Vega, Lorenzo Sindoni, Alessio Belenchia, Daniele Vernieri and Noemi Frusciante, for their keen remarks and stimulating questions; and Antonio Romano, for illuminating discussions. S.L. thanks Thanu Padmanabhan for insightful suggestions and shrewd considerations.

## Appendix A: First-order perturbations

In this Appendix, we collect a number of useful expressions for the differences, to the first order in  $\epsilon$ , between the geometric objects built out of two metrics  $g_{ab}$  and

---

vacuum solution. This fact, together with the formally different version of the same condition provided for scalar-tensor theories, viz.  $\nabla_d(\phi R_{abc}{}^d) = 0$ , strongly suggests that (2) may not be a universal law expressing the SEP/GWEP in a gauge-inspired representation of any extended theory of gravity.

---

<sup>21</sup> Taken as is, Eq. (2) is not verified already at the level of a single black hole-like solution for Gauß–Bonnet theory, let alone for any

$\bar{g}_{ab}$  connected via the relation (7). Thus, all the equations presented here hold only up to order  $\epsilon$ . First of all, we note that Eq. (7) implies

$$g^{ab} = \bar{g}^{ab} - \epsilon \gamma^{ab}, \quad (\text{A1})$$

where  $\bar{g}^{ab}$  is the inverse of  $\bar{g}_{ab}$ , and  $\gamma^{ab} := \bar{g}^{ac} \bar{g}^{bd} \gamma_{cd}$ .

To find the relation between the determinants of the metric coefficients, let us first expand  $g$  around the unperturbed metric  $\bar{g}_{ab}$ :

$$g = \bar{g} + \epsilon \frac{\partial g}{\partial \bar{g}_{ab}} \gamma_{ab}, \quad (\text{A2})$$

where the partial derivatives are evaluated at  $g_{ab} = \bar{g}_{ab}$ . Using the property  $\partial g / \partial g_{ab} = g g^{ab}$ , and defining  $\gamma := \bar{g}^{ab} \gamma_{ab}$ , we find the simple relation

$$g = \bar{g} (1 + \epsilon \gamma). \quad (\text{A3})$$

The Christoffel symbols  $\Gamma^a_{bc}$  and  $\bar{\Gamma}^a_{bc}$  of the metrics  $g_{ab}$  and  $\bar{g}_{ab}$ , respectively, are related as

$$\Gamma^a_{bc} = \bar{\Gamma}^a_{bc} + \epsilon \Xi^a_{bc}, \quad (\text{A4})$$

where  $\Xi^a_{bc} := \bar{g}^{ad} \Xi_{dbc}$ , and the tensor  $\Xi_{abc}$  is

$$\Xi_{abc} = \frac{1}{2} (\bar{\nabla}_b \gamma_{ca} + \bar{\nabla}_c \gamma_{ba} - \bar{\nabla}_a \gamma_{bc}). \quad (\text{A5})$$

This result can be easily obtained using the expression

$$\begin{aligned} \nabla_a g_{bc} &= \bar{\nabla}_a \bar{g}_{bc} - \epsilon \Xi^d_{ab} g_{dc} - \epsilon \Xi^d_{ac} g_{bd} \\ &= \bar{\nabla}_a \bar{g}_{bc} + \epsilon \bar{\nabla}_a \gamma_{bc} - \epsilon \Xi_{cab} - \epsilon \Xi_{bac}, \end{aligned} \quad (\text{A6})$$

which follows from Eq. (A4). Since the covariant derivatives  $\nabla_a$  and  $\bar{\nabla}_a$  are associated with  $g_{ab}$  and  $\bar{g}_{ab}$ , respectively, the compatibility condition for the Riemannian connection gives  $\nabla_a g_{bc} = \bar{\nabla}_a \bar{g}_{bc} = 0$ . Thus,

$$\Xi_{cab} + \Xi_{bac} = \bar{\nabla}_a \gamma_{bc}. \quad (\text{A7})$$

Equation (A5) is then obtained following the same steps by which one finds the usual expression for the Christoffel symbols in terms of partial derivatives of the metric.

The first-order difference  $\epsilon \mathcal{R}_{abc}^d = R_{abc}^d - \bar{R}_{abc}^d$  between the Riemann curvature tensors follows from Eq. (A4), and one has

$$\mathcal{R}_{abc}^d = \bar{\nabla}_b \Xi^d_{ac} - \bar{\nabla}_a \Xi^d_{bc}. \quad (\text{A8})$$

This implies, for the difference  $\epsilon \mathcal{R}_{ab} = R_{ab} - \bar{R}_{ab}$  between the Ricci tensors,

$$\mathcal{R}_{ab} = \mathcal{R}_{acb}^c = \bar{\nabla}_c \Xi^c_{ab} - \bar{\nabla}_a \Xi^c_{cb}; \quad (\text{A9})$$

and, for the difference  $\epsilon \mathcal{R} = R - \bar{R}$  between the curvature scalars  $R = g^{ab} R_{ab}$  and  $\bar{R} = \bar{g}^{ab} \bar{R}_{ab}$ ,

$$\mathcal{R} = \bar{g}^{ab} \bar{\nabla}_c \Xi^c_{ab} - \bar{g}^{ab} \bar{\nabla}_a \Xi^c_{cb} - \gamma^{ab} \bar{R}_{ab}, \quad (\text{A10})$$

where Eqs. (A1) and (A9) have been used.

Finally, for the difference  $\epsilon \mathcal{G}_{ab} = G_{ab} - \bar{G}_{ab}$  between the Einstein tensors we find, defining  $\Xi^{ab}_b := \bar{g}^{bc} \Xi^a_{bc}$ :

$$\begin{aligned} \mathcal{G}_{ab} &= \mathcal{R}_{ab} - \frac{1}{2} \bar{g}_{ab} \mathcal{R} - \frac{1}{2} \bar{R} \gamma_{ab} = \bar{\nabla}_c \Xi^c_{ab} - \bar{\nabla}_a \Xi^c_{cb} \\ &\quad - \frac{\bar{g}_{ab}}{2} \bar{\nabla}_c \Xi^{cd}_d + \frac{\bar{g}_{ab}}{2} \bar{\nabla}^c \Xi^e_{ec} + \frac{\bar{g}_{ab}}{2} \gamma^{cd} \bar{R}_{cd} - \frac{\gamma_{ab}}{2} \bar{R}. \end{aligned} \quad (\text{A11})$$

## Appendix B: Calculation of $\bar{\nabla}^b \mathcal{G}_{ab}$

The quantity  $\bar{\nabla}^b \mathcal{G}_{ab}$  intervenes frequently in the calculation of  $\bar{\nabla}^b \mathcal{E}_{ab}$  — noticeably, in Sec. IV A —, so we evaluate it here in full generality. We begin by substituting the expression (A5) for  $\Xi^a_{bc}$  into Eq. (A11), to obtain

$$\begin{aligned} \mathcal{G}_{ab} &= \frac{1}{2} \left[ (\bar{\nabla}^c \bar{\nabla}_a \gamma_{bc} + \bar{\nabla}^c \bar{\nabla}_b \gamma_{ac}) - \bar{\nabla}^c \bar{\nabla}_c \gamma_{ab} - \bar{\nabla}_a \bar{\nabla}_b \gamma \right. \\ &\quad \left. - \bar{g}_{ab} (\bar{\nabla}^c \bar{\nabla}^d \gamma_{cd} - \bar{\nabla}^c \bar{\nabla}_c \gamma) + \bar{g}_{ab} \gamma^{cd} \bar{R}_{cd} - \gamma_{ab} \bar{R} \right]. \end{aligned} \quad (\text{B1})$$

In a flat background spacetime (a situation common, for instance, in the study of gravitational radiation [6]), it is a straightforward exercise to show that  $\partial^b \mathcal{G}_{ab} = 0$ , the key point in the proof being a heavy use of the commutative property for partial derivatives. In the case of a non-flat background, on the other hand, switching covariant derivative operators  $\bar{\nabla}_a$  generates instances of the Riemann and Ricci tensors. The three terms in  $\bar{\nabla}^b \mathcal{G}_{ab}$  where this happens can be written, rearranging the indices and using the property  $\bar{\nabla}_a \bar{g}_{cd} = 0$ , as:

$$\begin{aligned} \bar{\nabla}_b \bar{\nabla}_c \bar{\nabla}_a \gamma^{bc} - \bar{\nabla}_a \bar{\nabla}_b \bar{\nabla}_c \gamma^{bc} &= \bar{R}_{abcd} \bar{\nabla}^d \gamma^{bc} \\ &\quad + 2 \bar{R}_{ab} \bar{\nabla}_c \gamma^{bc} + 2 \bar{\nabla}_b \bar{R}_{ac} \gamma^{bc} - \bar{\nabla}_a \bar{R}_{bc} \gamma^{bc}; \end{aligned} \quad (\text{B2})$$

$$\bar{\nabla}_b \bar{\nabla}_c \bar{\nabla}^b \gamma_a^c - \bar{\nabla}_c \bar{\nabla}_b \bar{\nabla}^b \gamma_a^c = -\bar{R}_{bcd a} \bar{\nabla}^b \gamma^{cd}; \quad (\text{B3})$$

$$\bar{\nabla}_a \bar{\nabla}_b \bar{\nabla}^b \gamma - \bar{\nabla}_b \bar{\nabla}_a \bar{\nabla}^b \gamma = -\bar{R}_{ab} \bar{\nabla}^b \gamma; \quad (\text{B4})$$

where in Eq. (B2) we have used the general identity (3) for the background quantities. Using these expressions, we find at the end

$$\begin{aligned} \bar{\nabla}^b \mathcal{G}_{ab} &= \frac{1}{2} (2 \bar{R}_{ab} \bar{\nabla}_c \gamma^{bc} + 2 \gamma^{bc} \bar{\nabla}_b \bar{R}_{ac} - \bar{R}_{ab} \bar{\nabla}^b \gamma \\ &\quad + \bar{R}_{bc} \bar{\nabla}_a \gamma^{bc} - \bar{R} \bar{\nabla}^b \gamma_{ab} - \gamma_{ab} \bar{\nabla}^b \bar{R}). \end{aligned} \quad (\text{B5})$$

- 
- [1] C. M. Will, *Theory and Experiment in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, England, 1993).
  - [2] C. M. Will, “The confrontation between general relativity and experiment”, *Living Rev. Relativity* **9**, 3 (2006).
  - [3] E. Di Casola, S. Liberati, and S. Sonego, “Nonequivalence of equivalence principles”, submitted; E-print gr-qc/1310.7426.
  - [4] F. Ravndal, “Scalar gravitation and extra dimensions”, in *Proceedings of the Gunnar Nordström Symposium on Theoretical Physics, Helsinki, 2003*, edited by C. Cronström and C. Montonen, *Commentationes Physico-Matematicæ* **166** (2004), pp. 151–164; E-print gr-qc/0405030.
  - [5] D. Giulini, “What is (not) wrong with scalar gravity?”, *Stud. Hist. Phil. Mod. Phys.* **39**, 154–180 (2008); E-print gr-qc/0611100.
  - [6] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
  - [7] E. Poisson, A. Pound, and I. Vega, “The motion of point particles in curved spacetime”, *Living Rev. Relativity* **14**, 7 (2011); E-print gr-qc/1102.0529.
  - [8] E. Poisson, “The gravitational self-force”, in *General Relativity and Gravitation: Proceedings of the 17th International Conference*, edited by P. Florides, B. Nolan, and A. Ottewill (Singapore, World Scientific, 2005), pp. 119–141; E-print gr-qc/0410127.
  - [9] A. Papapetrou, “Spinning test-particles in general relativity. I”, *Proc. R. Soc. London A* **209**, 248–258 (1951).
  - [10] Y. Mino, M. Sasaki, and T. Tanaka, “Gravitational radiation reaction to a particle motion”, *Phys. Rev. D* **55**, 3457–3476 (1997); E-print gr-qc/9606018.
  - [11] T. C. Quinn and R. M. Wald, “Axiomatic approach to electromagnetic and gravitational radiation reaction of particles in curved spacetime”, *Phys. Rev. D* **56**, 3381–3394 (1997); E-print gr-qc/9610053.
  - [12] J. D. Bekenstein, “The modified Newtonian dynamics—MOND and its implications for new physics”, *Contemp. Phys.* **47**, 387–403 (2006); E-print astro-ph/0701848.
  - [13] K. Nordvedt, “Equivalence principle for massive bodies. II. Theory”, *Phys. Rev.* **169**, 1017–1025 (1968).
  - [14] J.-M. Gérard, “The strong equivalence principle from a gravitational gauge structure”, *Class. Quantum Grav.* **24**, 1867–1877 (2007); E-print gr-qc/0607019.
  - [15] J.-M. Gérard, “Further issues in fundamental interactions”, in *Proceedings of the 2008 European School of High-Energy Physics*, edited by N. Ellis and R. Fleischer (CERN, Geneva, 2009), pp. 281–314; E-print hep-ph/0811.0540.
  - [16] R. P. Geroch and P. S. Jang, “Motion of a body in general relativity”, *J. Math. Phys. (N.Y.)* **16**, 65–67 (1975).
  - [17] J. Ehlers and R. P. Geroch, “Equation of motion of small bodies in relativity”, *Ann. Phys. (N.Y.)* **309**, 232–236 (2004); E-print gr-qc/0309074.
  - [18] D. B. Malament, *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory* (University of Chicago Press, Chicago, 2012), pp. 146–149.
  - [19] N. Deruelle, “Nordström’s scalar theory of gravity and the equivalence principle”, *Gen. Relativ. Gravit.* **43**, 3337–3354 (2011); E-print gr-qc/1104.4608.
  - [20] R. Infeld and A. Schild, “On the motion of test particles in general relativity”, *Rev. Mod. Phys.* **21**, 408–413 (1949).
  - [21] J. W. York, Jr., “Role of conformal three-geometry in the dynamics of gravitation”, *Phys. Rev. Lett.* **28**, 1082–1085 (1972).
  - [22] G. W. Gibbons and S. W. Hawking, “Action integrals and partition functions in quantum gravity”, *Phys. Rev. D* **15**, 2752–2756 (1977).
  - [23] E. Dyer and K. Hinterbichler, “Boundary terms, variational principles, and higher derivative modified gravity”, *Phys. Rev. D* **79**, 024028 (2009); E-print gr-qc/0809.4033.
  - [24] V. Faraoni, *Cosmology in Scalar-Tensor Gravity* (Kluwer, Dordrecht, 2004).
  - [25] T. P. Sotiriou and V. Faraoni, “ $f(R)$  theories of gravity”, *Rev. Mod. Phys.* **82**, 451–497 (2010); E-print gr-qc/0805.1726.
  - [26] A. Guarnizo, L. Castañeda, and J. M. Tejeiro, “Boundary term in metric  $f(R)$  gravity: field equations in the metric formalism”, *Gen. Relativ. Gravit.* **42**, 2713–2728 (2010); E-print gr-qc/1002.0617.
  - [27] T. P. Sotiriou, S. Liberati, and V. Faraoni, “Theory of gravitation theories: A no-progress report”, *Int. J. Mod. Phys. D* **17**, 399–423 (2008); E-print gr-qc/0707.2748.
  - [28] H.-J. Schmidt, “Fourth-order gravity and conformal transformations”, *Class. Quantum Grav.* **6**, 557–559 (1989).
  - [29] S. Gottlöber, H.-J. Schmidt, and A. A. Starobinsky, “Sixth-order gravity and conformal transformations”, *Class. Quantum Grav.* **7**, 893–900 (1990).
  - [30] A. Baykal and Ö. Delice, “Multi-scalar-tensor equivalents for modified gravitational actions”, *Phys. Rev. D* **88**, 084041 (2013); E-print gr-qc/1308.6106.
  - [31] D. Lovelock, “The Einstein tensor and its generalizations”, *J. Math. Phys.* **12**, 498–501 (1971).
  - [32] T. Padmanabhan, *Gravitation: Foundations and Frontiers* (Cambridge University Press, Cambridge, 2010).
  - [33] T. S. Bunch, “Surface terms in higher derivative gravity”, *J. Phys. A* **14**, L139–L143 (1981).
  - [34] R. C. Myers, “Higher-derivative gravity, surface terms, and string theory”, *Phys. Rev. D* **36**, 392–396 (1987).
  - [35] Y. Sendouda, N. Deruelle, M. Sasaki, and D. Yamauchi, “Higher curvature theories of gravity in the ADM canonical formalism”, *Int. J. Mod. Phys. Conf. Ser.* **01**, 297–302 (2011).
  - [36] A. Mukhopadhyay and T. Padmanabhan, “Holography of gravitational action functionals”, *Phys. Rev. D* **74**, 124023 (2006); E-print hep-th/0608120.
  - [37] R. Brustein and A. J. M. Medved, “Lovelock gravity is equivalent to Einstein gravity coupled to form fields”, *Phys. Rev. D* **88**, 064010 (2013); E-print hep-th/1212.0625.
  - [38] N. Deruelle, J. Katz, and S. Ogushi, “Conserved charges in Einstein Gauss-Bonnet theory”, *Class. Quantum Grav.* **21**, 1971–1985 (2004); E-print gr-qc/0310098.
  - [39] Y. Choquet-Bruhat, “Gravitation with Gauss Bonnet terms”, in Y. Choquet-Bruhat, *General Relativity and the Einstein Equations* (Oxford University Press, Oxford,



- 2009), pp. 689–708.
- [40] S. Golod and T. Piran, “Choptuik’s critical phenomenon in Einstein–Gauss–Bonnet gravity”, *Phys. Rev. D* **85**, 104015 (2012); E-print gr-qc/1201.6384.
- [41] P. A. R. Ade *et al.* [Planck Collaboration], “Planck 2013 results. XVI. Cosmological parameters”, E-print astro-ph.CO/1303.5076.